

BRIDGING COMPRESSION TO WAVELET THRESHOLDING AS A DENOISING METHOD

*S. Grace Chang*¹

*Bin Yu*²

*Martin Vetterli*¹

¹Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720, USA

²Department of Statistics
University of California, Berkeley, CA 94720, USA
grchang,martin@eecs.berkeley.edu, binyu@stat.berkeley.edu

ABSTRACT

Some past work has suggested that lossy compression can be a good denoising tool. Building on this theme, we make the connection that quantization of transform coefficients approximates the operation of Donoho-Johnstone's wavelet thresholding, to conclude that compression (via coefficient quantization) is appropriate for filtering noise from signal. The method of quantization is scale adaptive and is facilitated by a criterion similar to Rissanen's minimum description length principle. Results show that a small number of quantization levels achieves almost the same performance of full precision thresholding, suggesting that denoising is mainly due to the zero-zone and that the full precision of the threshold coefficients is of secondary importance.

1. INTRODUCTION

The purpose of this paper is to explain why lossy compression can be appropriate for signal denoising. More specifically, we wish to show that quantization (a common step in compression) of transform coefficients achieves denoising by posing quantization as an approximation to an effective denoising method called *wavelet thresholding* [4].

Donoho and Johnstone [4] have shown the effectiveness of using hard-thresholding and soft-thresholding for the removal of additive noise of normal distribution. The thresholding methods compare the input to a given threshold and set it to zero if its magnitude is less than the threshold. It essentially creates a region around zero where the coefficients are considered to be negligible, called the "zero-zone" (or "dead-zone") in the data compression community. Outside of this region, however, the thresholded coefficients are kept to full precision.

Recognizing that in a typical transform domain compression scheme, the coefficients are quantized with a zero-zone for negligible coefficient values, we show that an appropriate quantization (and hence compression) achieves denoising because it is an approximation to the thresholding function (see Figure 1). We hypothesize that the effectiveness of denoising is mainly due to the zero-zone, and that the full precision of the thresholded coefficients is of secondary importance. Thus, a comparable level of denoising performance can be achieved by quantizing the coefficients with a zero-zone and a few number of quantization levels outside the zero-zone. The selection of the number of quantization levels is facilitated by a criterion similar to Rissanen's min-

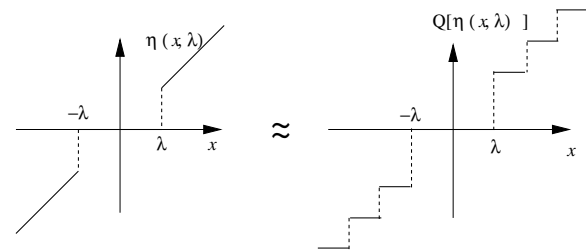


Figure 1. The hard-thresholding function can be approximated by quantization with a zero-zone.

imum description length (MDL) [9] and it is done in a scale adaptive manner.

The idea of using a lossy compression scheme for denoising have been proposed by several works. Saito [10] combined the ideas of wavelet thresholding and the MDL principle to achieve simultaneous denoising and compression, where the compression results from the fact that only a subset of the coefficients are nonzero. However, this work does not address the issue that in a practical compression framework, the parameters are usually quantized and not kept to full precision (i.e. machine precision). Moulin [6, 7] has also shown some interesting results connecting thresholding and MDL for signal estimation.

Another work of interest is by Natarajan [8], who proposed to remove noise via a lossy compression whose distortion is set at the noise strength. This work elucidates the idea that lossy compression is good for noise removal, and showed the error residual to be approximately proportional to $(s/N)^{1/2}$, where N is the number of data samples, and s is the resulting number of bits when the signal is compressed at distortion equal to the noise strength. The framework is for any general compression algorithm. Here we pinpoint the effectiveness of denoising to be derived mainly from thresholding insignificant transform coefficients.

The organization of the paper is as follows. In Section 2, the background material concerning wavelet thresholding is briefly introduced. In Section 3, we discuss the issue of quantizing the coefficients with an MDL-like criterion, and some preliminary analysis is presented in Section 4. Experimental results are presented in Section 5, with a discussion on application to images. Lastly, Section 6 summarizes the main points and a list of ongoing and future work.

2. WAVELET THRESHOLDING FOR DENOISING

In this section, we introduce briefly the framework of Donoho and Johnstone, and refer the reader to [4] and related publications for more detail. There are many references for wavelet theory, such as [2, 3, 11], and we omit the details here.

2.1. Definitions and Notations

Suppose we have an unknown function $f(t)$ defined on $t \in [0, 1]$, and we sample it on a uniform spacing of $1/N$. Further suppose that these samples have been corrupted by additive noise and we observe

$$y_i = f(t_i) + v_i, \quad i = 1, 2, \dots, N \quad (1)$$

where $t_i = (i-1)/N$, v_i are *iid* $N(0, \sigma^2)$ for some known σ^2 , and $N = 2^J$. Let $f = \{f(t_i)\}_{i=1}^N$ be the vector of the sampled points $f(t_i)$. The goal is to obtain an estimate \hat{f} with a small risk $E\|\hat{f} - f\|_2^2$.

To remove noise and obtain an estimate \hat{f} , Donoho and Johnstone [4] proposed to perform coefficient thresholding in the wavelet domain and demonstrated asymptotic minimax optimality of such an operation. Let $y = \{y_i\}_{i=1}^N$ denote the vector of observations, and $Y = Wy$ the vector of N wavelet coefficients of y , where W is the matrix implementing the orthogonal wavelet transform (periodized at the boundary). Since the transform is orthogonal, the inverse transform is $y = W^T Y$.

It is sometimes convenient to index Y as

$$Y_{j,k} = (Wy)_{j,k}, \quad j = 0, \dots, J-1, \quad k = 1, \dots, 2^j,$$

and the remaining element $Y_{-1,1}$. The index j indicates the resolution level (large j for fine scale, and small j for coarse scale). Practically, one considers $j = L, \dots, J-1$, where the parameter $L > 0$ denotes the coarsest resolution level j considered in the wavelet transform. The low resolution residual is indexed by $j = 0, k = 1, \dots, 2^L$. For simpler notations, in the subsequent text these coefficients are referenced with two indices when the scale j is considered, and a single index otherwise; the distinction will be clear from the context.

Similarly, let $F = Wf$ be the wavelet transform of the signal $f = \{f(t_i)\}_{i=1}^N$, and $V = Wv$ the wavelet transform of the noise sequence $v = \{v_i\}_{i=1}^N$. Since v_i are white noise, $V_{j,k}$ are also *iid* white noise $N(0, \sigma^2)$.

The signal estimation is done in the wavelet domain by thresholding the coefficients. Define η_h to be the hard-thresholding function $\eta_h(x, \lambda) = x \cdot \mathbf{1}\{|x| > \lambda\}$, which keeps the coefficient if its magnitude is larger than λ . Another function of interest is the soft-thresholding function $\eta_s(x, \lambda) = \text{sgn}(x)(|x| - \lambda)_+$, which shrinks the coefficient towards 0 by λ .

The proposed estimator is constructed by hard-thresholding the wavelet coefficients using the universal threshold $\lambda = \lambda_U = \sigma\sqrt{2\log N}$ and transform them back to the space domain. That is,

$$\hat{F}_{j,k} = \eta_h(Y_{j,k}, \lambda_U),$$

and

$$\hat{f} = W^T \hat{F}$$

is the estimated signal. Note that the thresholding operation is performed only on resolution levels $j = L, \dots, J-1$, and the $j = 0$ low resolution coefficients are kept without thresholding.

2.2. Theoretical properties

This estimator has been shown to have some desirable minimax properties, and the subsequent discussion is set in the wavelet domain. Suppose the observations are $Y_i = F_i + V_i, i = 1, \dots, N$, where V_i are *iid* $N(0, \sigma^2)$ for some known σ^2 and F_i is the deterministic signal to be estimated.

For many regular functions, only a few wavelet coefficients are substantially different from zero, thus heuristically, one could argue that a good estimate could be obtained by setting to zero the insignificant coefficients, which are likely due to noise. Hence, consider among all *diagonal projection* estimators of the form $\hat{F}_i^{DP} = \delta_i Y_i$, $\delta_i = 0$ or 1 . When the error V_i is *iid* $N(0, \sigma^2)$, the ideal diagonal projection estimator is obtained by setting $\delta_i = \mathbf{1}\{|F_i| > \sigma\}$, and the associated risk is

$$E\|\hat{F}_i^{DP} - F\|^2 = \sum_i \min(F_i^2, \sigma^2).$$

This ideal estimator cannot be used since it requires F to be known, however it serves as a useful benchmark. More specifically, the benchmark risk is $\sigma^2 + \sum_i \min(F_i^2, \sigma^2)$, and for large N , all but the sparsest signals will be essentially compared against $E\|\hat{F}_i^{DP} - F\|^2$.

For the estimate $\hat{F}_i = \eta_h(Y_i, \lambda)$ with $\lambda = \sigma\sqrt{2\log N}$, [4] showed it to satisfy

$$E\|\hat{F} - F\|^2 \leq (1 + 2\log N)(\sigma^2 + \sum_i \min(F_i^2, \sigma^2)).$$

That is, the error is within $2\log N$ factor of the ideal risk $R(\hat{F}_i^{DP}, F)$. Furthermore, for all estimator of the form $\hat{F}_i = \theta(Y_i)$ for some function $\theta(\cdot)$,

$$\frac{1}{2\log N} \inf_{\hat{F}} \sup_{F \in \mathbb{R}^N} \frac{E\|\hat{F} - F\|^2}{\sigma^2 + \sum_{i=1}^N \min(F_i^2, \sigma^2)} \rightarrow 1 \text{ as } N \rightarrow \infty.$$

Hence, the best estimator \hat{F} yields a maximum risk over all $F \in \mathbb{R}^N$ that grows as $2\log N$ of the ideal risk. Thus, the hard-threshold estimate is asymptotically optimal in the *minimax* sense (minimizing the maximum error).

3. QUANTIZATION USING MDL SELECTION

The purpose of this work is to show that quantization could be a good approximation of the thresholding function, and thus compression (via quantization) can be used for denoising. This section describes the method of quantization using an MDL criterion, much like in [10, 6, 7], but restricting the estimate to belong to the set of quantized signals defined in the subsequent text.

The boundary of the quantization zero-zone is set to the threshold λ . The next question is how to quantize the coefficients outside of the zero-zone; that is, how many quantization levels and what type of reconstruction values (e.g.

uniform, centroid, Lloyd-Max). An MDL-like criterion is used for this decision.

For a given set of observations, the MDL criterion is useful for choosing a reasonable statistical model which gives the shortest description length. It does this by choosing the model which minimizes the total codelength of a two-part encoding consisting of the data (based on the chosen model) and of the model parameters. The idea is that the chosen model should establish a compromise between fitting the data well and having low complexity (i.e. having a simple representation or a reasonable number of parameters).

More specifically, given the set of noisy transform coefficients $Y = F + V$, the framework is to code Y given the model F (which, of course, needs to be estimated). The MDL principle chooses F which minimizes the two-part codelength

$$L_{\text{total}}(Y, F) = L(Y|F) + L(F),$$

where $L(Y|F)$ is the codelength for Y based on the model F , and $L(F)$ is the codelength for F . If it is possible to associate a probability distribution p , then one can use the idealized codelength $-\log_2 p$. In this case where we assume the noise V_i to be *iid* $N(0, \sigma^2)$, the first term becomes

$$\begin{aligned} L(Y|F) &= -\log_2 p(Y|F) \\ &= -\log_2 \left(\frac{1}{\sqrt{2\pi\sigma^{2N}}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^N (Y_i - F_i)^2 \right\} \right) \\ &= \frac{1}{2(\ln 2)\sigma^2} \sum_{i=1}^N (Y_i - F_i)^2 + \frac{1}{2} \log_2(2\pi\sigma^{2N}) \end{aligned} \quad (2)$$

The idea is to minimize the criterion over estimates of F residing in the set of quantized signals, whose description will lead to defining $L(F)$ as well.

First assume that the coefficients are quantized uniformly with equal width bins and the reconstruction values are the mid-point of the bins. That is, we partition K equal bins between the threshold λ and $|Y|_{\text{max}}$, the largest coefficient magnitude, and have K bins on each of the positive and negative side. The coefficients whose absolute values are smaller than λ are set to 0. Along with the zero-zone, this yields $2K + 1$ bins. In the coding parlance, there is an *overhead* of coding the values K , λ , and $|Y|_{\text{max}}$. However, since this is the same for all K , it is immaterial to the minimization. Each of the N coefficient needs $\log_2(2K + 1)$ bits to denote which bin it resides in. Hence, assuming uniform indexing,

$$L(F) = N \log_2(2K + 1) \quad (3)$$

Let $Q[x, K]$ denote the operation of quantizing the input x to $2K + 1$ levels as described above, and let $\hat{F}_i^Q = Q[\eta_h(Y_i, \lambda), K]$ denote the estimate obtained by first hard-thresholding followed by quantization (note that for the hard-thresholding case, this is the same as if we quantize on Y_i directly, but the notation is kept as such to incorporate the case of soft-thresholding). Then one chooses the

estimate \hat{F}^Q with the associated K to minimize the criterion (combining (2) and (3))

$$MDLQ = \frac{1}{2(\ln 2)\sigma^2} \sum_{i=1}^N (Y_i - \hat{F}_i^Q)^2 + N \log_2(2K + 1) \quad (4)$$

The time-domain estimate is then taken to be $W^T \hat{F}^Q$.

Note that the quantized estimate naturally could do worse than the unquantized threshold estimate. However, it is not our goal to achieve better than the unquantized estimate, but rather to establish a connection between compression (via quantization) and thresholding to show that compression is indeed a good method for noise removal.

A natural extension is to have different number of levels for each resolution scale, since the low resolution coefficients are more important and usually allocated more bits. Thus for each scale $j = L, \dots, J - 1$, we find $\hat{F}_{j,k}^Q$, $k = 1, \dots, 2^j$ and the associated K_j which minimizes

$$MDLQ_j = \frac{1}{2(\ln 2)\sigma^2} \sum_{k=1}^{2^j} (Y_{j,k} - \hat{F}_{j,k}^Q)^2 + 2^j \log_2(2K_j + 1). \quad (5)$$

The scale $j = 0$ coefficients are usually coded with high precision because of its importance. We experimented with keeping its full precision or using (5) and found that the criterion does indeed yield a large K_j value for the low resolution level $j = 0$, as should be in practical coding.

4. THEORETICAL BOUNDS

We have some preliminary theoretical bounds, but they are too loose to be of practical use. The goal is to compare the risk $E\|\hat{F}^Q - F\|^2$ to $E\|\hat{F} - F\|^2$, and show that the former achieves within a constant (which may depend on the quantization binwidth) of the latter. Since each Y_i is *iid* $N(F_i, \sigma^2)$, it suffices to examine the risk for one variable, $E(\hat{F}_i^Q - F_i)^2$. Assume uniform quantization with mid-point reconstruction. Let Δ be the binwidth of the quantization bin, $a_k = \lambda + k\Delta$, $k = 0, 1, \dots$, be the bin boundaries (similarly for the negative side), $a_k + \Delta/2$ be the reconstructed value in bin k , and $[a_k, a_k + \Delta]$ be the k -th quantization bin. Consider the ratio

$$\frac{\int_{a_k}^{a_k+\Delta} (a_k + \Delta/2 - F_i)^2 N_x(F_i, \sigma^2) dx}{\int_{a_k}^{a_k+\Delta} (x - F_i)^2 N_x(F_i, \sigma^2) dx} \quad (6)$$

where the subscript x in $N_x(F_i, \sigma^2)$ signifies that x is the Gaussian distributed random variable. For any value of F_i , λ , Δ , and any quantization bin, it can be shown that (6) has an upper bound $C(\Delta) = \max(5, (27/28)e^{(3\Delta^2)/(2\sigma^2)})$. Hence,

$$E(\hat{F}_i^Q - F_i)^2 \leq C(\Delta) E(\hat{F}_i - F_i)^2.$$

Numerical calculation shows that the bound is much less (see Figure 2), and is approximately $e^{(\Delta/\sigma)/3}$ for small Δ/σ and $(\frac{\Delta}{\sigma})^2/4$ for larger Δ/σ . The plot in Figure 2 plots $r(\Delta/\sigma)$ as a function of Δ/σ , where

$$r\left(\frac{\Delta}{\sigma}\right) = \max_{F_i} \max_{\lambda} \frac{E(\hat{F}_i^Q - F_i)^2}{E(\hat{F}_i - F_i)^2}.$$

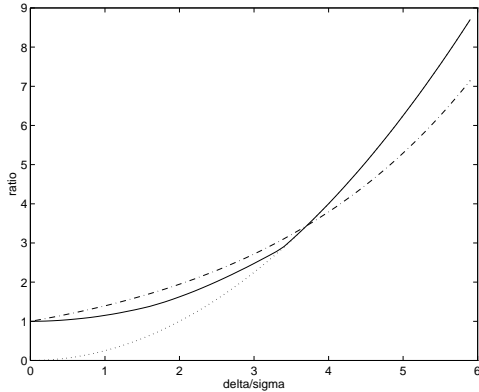


Figure 2. Plot the risk ratio $r(\frac{\Delta}{\sigma})$ (—), compared with $(\frac{\Delta}{\sigma})^2/4$ (···), and $e^{(\Delta/\sigma)^3}$ (- · - ·).

Data	(a)	(b)	(c)	(d)
Full prec.	.299	.320	.066	.115
Quantized	.445	.659	.086	.207

Table 1. Comparing the MSE of the full precision hard-thresholded estimate and the MSE of the quantized estimate, for the synthetic data set.

Further investigation is underway to find a tighter analytical bound.

5. EXPERIMENTAL RESULTS

5.1. Synthetic 1-D Data

We compare the quantized estimate against the full precision threshold estimate using the test data in [4]. Figure 3 shows the original data set, and Figure 4 shows the noisy observation. Each data sequence has 2048 samples, scaled such that $\|f\|/\|v\| = 7$ and the noise v_i is *iid* $N(0, 1)$. The chosen wavelet is Daubechies' Symmlet wavelet with parameter 8 (7 vanishing moments) [3], and the lowest resolution considered is $L = 6$ (i.e. 5 scales of decomposition). The denoised estimates using hard-thresholding with $\lambda_U = \sigma\sqrt{2\log N}$ are shown in Figure 5 and the quantized estimates using $MDLQ_j$ in (5) with uniform quantization are illustrated in Figure 6.

To gather some statistical data, we run 100 simulations. The MSE are shown in Table 1, and the average numbers of quantization levels K_j are shown in Table 2. Visually the reconstructed waveforms are very similar, even if the values of K_j are quite small. A substantial amount of the distortion is due to quantizing the low resolution coefficients, suggesting a more careful treatment. For (a) and (b), quantization actually makes the flat regions less wiggly. The results show that the $MDLQ_j$ criterion allocates more levels in the coarser, more important levels, as would be the case in a practical subband coding situation. The MSE of the quantized estimate is comparable to the full-precision case, although for case (b), it is about twice as large, due to the quantization of large coefficients (in the regions of sharp transition).

scale j	10	9	8	7	6	0
(a)	1.00	1.00	1.00	1.99	4.70	16.72
(b)	1.00	1.00	2.72	6.60	13.33	22.78
(c)	0.09	0.08	0.36	0.83	0.99	17.91
(d)	0.12	0.90	1.00	1.98	2.92	15.69

Table 2. The values of K_j , $j = 6, \dots, 10$ and $j = 0$ (low resolution) for each signal, averaged over 100 runs. A smaller j corresponds to a coarser scale, and a larger j corresponds to a finer scale.

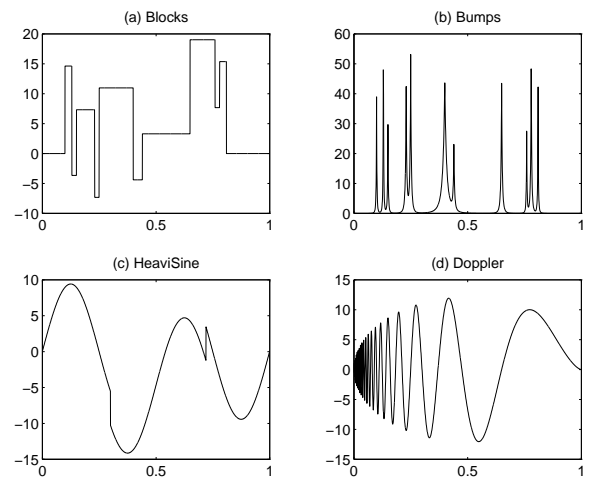


Figure 3. Original data.

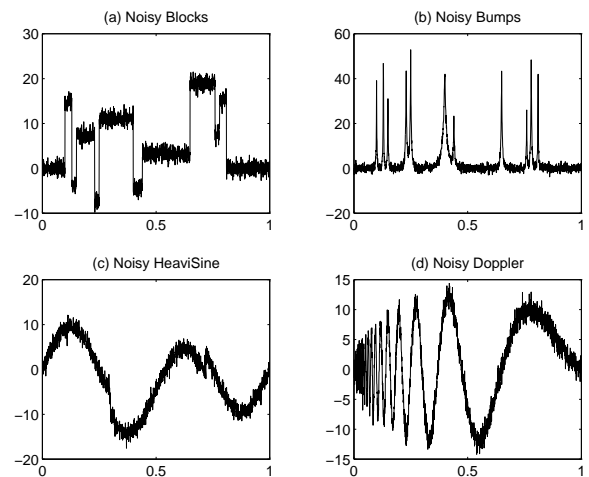


Figure 4. Data with additive *iid* $N(0, 1)$ noise.

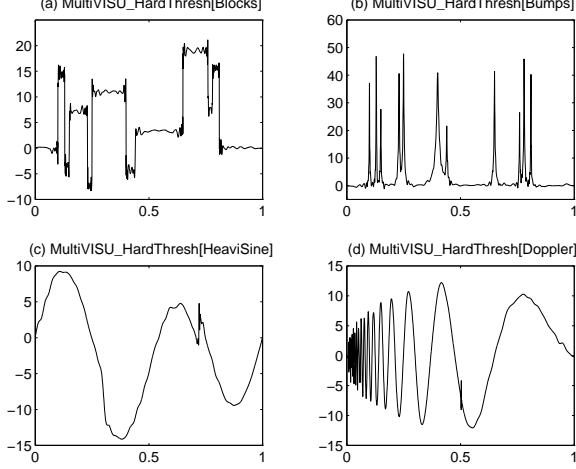


Figure 5. Estimates generated from hard-thresholding with fixed threshold λ_U .

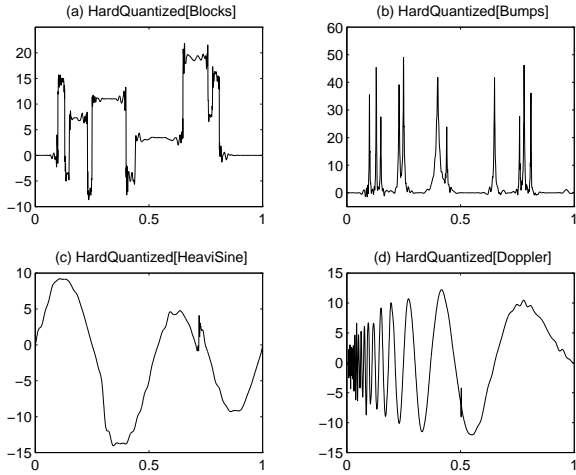


Figure 6. Estimates generated from hard-thresholding with threshold λ_U , followed by quantization according to the $MDLQ_j$ criterion.

5.2. Application to Images

To apply $MDLQ$ to images, several changes are made for better performance. The details are described in [1], and will be summarized here. The soft-thresholding function is used instead, because it yields smoother images, while hard-thresholding tends to produce unpleasant blips. Furthermore, the theoretical thresholds such as λ_U and other variations do not work for images and do not make sense (due to the dependence on the sample size N). Thus, the generalized cross-validation (GCV) method in [5] is used in each subband to find appropriate (and adaptive) thresholds. This criterion is

$$GCV(\lambda) = \frac{\frac{1}{N} \sum (Y_i - \eta_s(Y_i, \lambda))^2}{[\# \text{zero coefficients}]/N]^2}, \quad (7)$$

the optimal threshold λ is one that minimized (7), and N is the total number of pixels in the subband. This criterion can be shown to approximate the risk asymptotically, and

Data Set	Lena	Barbara
PSNR full prec.	33.09	30.78
PSNR quantized	31.62	28.95

Table 3. Comparing the PSNR (dB) of the full precision soft-thresholded estimate and the PSNR of the quantized estimate.

Orientation	Scale			
	8	7	6	5
HH	1.0	3.0	6.6	14.6
HL	2.4	7.0	10.8	14.8
LH	1.2	5.4	7.8	13.4
LL	24.6			

Table 4. Averaged over 5 runs, the value of K_j for the different orientation and scale components of Lena.

its derivation is found in [5].

The MDL criterion and quantization also need to be changed. It is well known that for a natural image, each subband (except the lowest) of its wavelet transform tends to be Laplacian distributed. Assuming that $Y_i, i = 1, \dots, N$ are *iid* Laplacian distributed, or equivalently, that $|Y_i|, i = 1, \dots, N$ are *iid* exponentially distributed, defined as $p(x) = \alpha e^{-\alpha x}, x \geq 0$. (Note: this is not to be confused with Y_i conditioned on knowing F_i is Gaussian distributed.) The parameter α can be estimated from the soft-thresholded coefficients, $\hat{F}_i = \eta_s(Y_i, \lambda)$, and the maximum likelihood estimate is $\hat{\alpha} = M/(\sum_i |\hat{F}_i|)$, where M is the number of nonzero coefficients. The bins are equally divided, and the reconstructed values are taken to be the centroids in each bin, calculated based on the estimated Laplacian parameter.

The formulation of $L(F)$ is also changed to assume a more efficient encoding. To indicate the locations of the nonzero coefficients, $M \log 2(N)$ bits are needed, but this number is constant once a threshold is chosen. A good estimate of the required bits for the M nonzero coefficients is the zeroth order entropy $H(K) = -\sum_k m_k \log_2(m_k/M)$, where m_k is the number of values in bin k , and the sum is over all $2K$ quantization bins. Thus $L(F) = H(K)$. As in the 1-D case, the $MDLQ$ criterion is minimized for each subband independently. The lowest resolution band is quantized assuming uniform distribution, while the other subbands assume Laplacian distribution.

This method is used on several test images, with noise *iid* $N(0, \sigma = 10)$ (which yields a PSNR of 28.13 dB), for 5 runs. Again, the same Symmlet wavelet is used, with separable horizontal and vertical filtering. Table 3 compares the PSNR between the full precision soft-thresholded estimate and the quantized. Table 4 shows the number of quantization levels used. Figure 7 shows the noisy image, the thresholded image and the quantized image.



Figure 7. (a) Noisy image with noise $N(0, \sigma = 10)$. (b) Thresholded image and (c) quantized image.

6. DISCUSSION AND CONCLUSION

In this paper we have demonstrated the connection between quantization and thresholding, thus further validating the claim of using compression as a denoising method. In particular, we showed that the zero-zone from quantizing is the main contributor to denoising and that outside of this region, the coefficients can be quantized to a lower precision. Furthermore, the results suggest that the proposed scheme may be appropriate for subband coding, in the decision of bit allocation, the zero-zone, and the manner of quantization.

Further work involves deriving other variations of the *MDLQ* criterion, and also addressing the issue of when σ is unknown. In the analytical aspect, more work is needed to find a tighter error bound on the quantized estimate, to show that it achieves within a reasonable constant of the threshold estimate. Another topic of interest is to develop a different thresholding framework, where the threshold does not depend on the value of N , but just on σ , since this is more intuitive for signal processing applications.

REFERENCES

- [1] S.G. Chang, B. Yu, and M. Vetterli, "Lossy Compression for Noise Removal," submitted to *Int. Conf. Image Processing*, 1997.
- [2] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1992.
- [3] I. Daubechies, "Orthonormal bases of compactly supported wavelets II: Variations on a theme," *SIAM J. Math. Anal.*, vol 24, pp. 499-519, 1993.
- [4] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol 81, pp. 425-455, 1994.
- [5] M. Jansen, M. Malfait, A. Bultheel, "Generalized cross validation for wavelet thresholding," preprint, Dec. 1995.
- [6] P. Moulin, "Model selection criteria and the orthogonal series method for function estimation," *Proc. IEEE Symp. on Info Thy.*, Whistler, B.C., p. 252, Sep. 1995.
- [7] P. Moulin, "Signal estimation using adapted tree-structured bases and the MDL principle," *Proc. IEEE Time-Frequency and Time-Scale Analysis*, pp. 141-143, June 1996.
- [8] B.K. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Trans. on Signal Processing*, vol. 43, no. 11, pp. 2595-2605, November 1995.
- [9] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, 1989.
- [10] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelets in Geophysics*, Eds. E. Foufoula-Georgiou and P. Kumar, pp. 299-324, Academic Press, 1994.
- [11] M. Vetterli, J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.